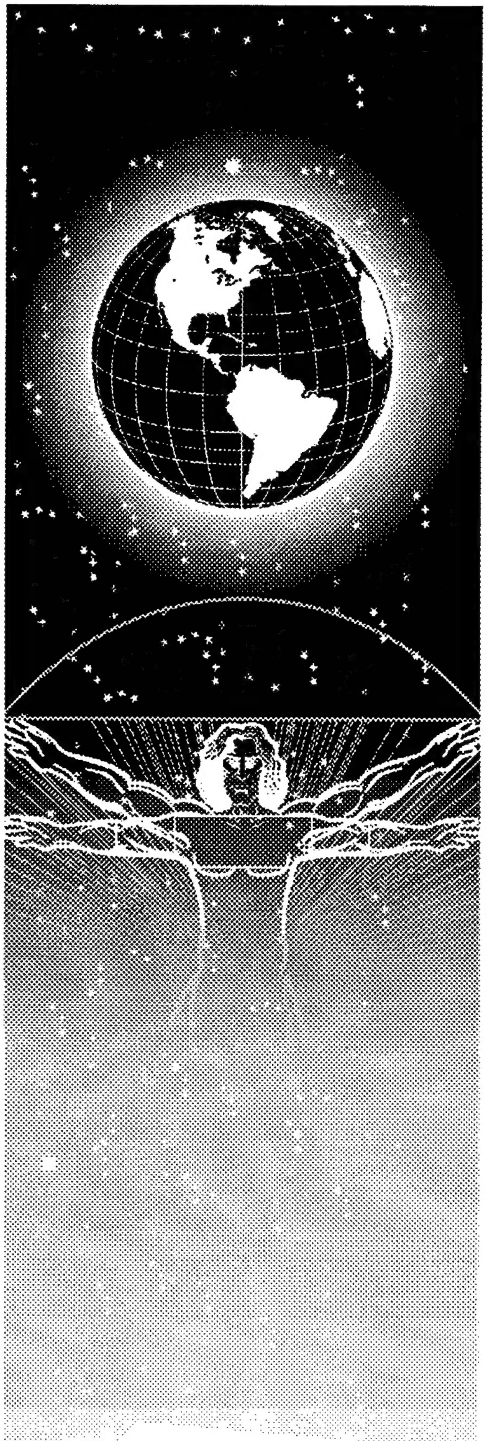
AL/CF-TR-1997-0017



UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

OPERATOR WORKLOAD IN THE F-15E: A COMPARISON OF TAWL AND MICRO SAINT COMPUTER SIMULATIONS (U)

Judi E. See

LOGICON TECHNICAL SERVICES, INC.
P.O. BOX 317258
DAYTON, OH 45437-7258

Michael A. Vidulich

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION
WRIGHT-PATTERSON AFB, OH 45433-7022

DTIC QUALITY INSPECTED 2

JANUARY 1997

INTERIM REPORT FOR THE PERIOD APRIL 1995 TO DECEMBER 1996

19970730 082

Approved for public release; distribution is unlimited.

Crew Systems Directorate
Human Engineering Division
2255 H Street
Wright-Patterson AFB OH 45433-7022

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from Armstrong Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, Virginia 22060-6218

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1997-0017

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Kenneth R. Boff

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
January 1997

3. REPORT TYPE AND DATES COVERED
Interim, Apr 1995-Dec 1996

4. TITLE AND SUBTITLE
Operator Workload in the F-15E: A Comparison of TAWL and Micro Saint Computer Simulations (U)

5. FUNDING NUMBERS
F41624-94-C-6007
PE 62202F
PR 7184
TA 14
WU 25

6. AUTHOR(S)
*Judi E. See
Michael A. Vidulich

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
*Logicon Technical Services, Inc.
P.O. Box 317258
Dayton OH 45437-7258

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
Human Engineering Division
Human Systems Center
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/CF-TR-1997-0017

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
The mental workload experienced by the crewmember occupying the back seat of the F-15E during a target acquisition mission was simulated via two computer modeling tools: Task Analysis/ Workload (TAWL) and the microcomputer version of Systems Analysis of Integrated Networks of Tasks (Micro Saint). The primary objectives were to evaluate the similarity of the two modeling tools and compare their relative ease of use. The scenario consisted of a ten-task target acquisition mission whose goal was to detect and destroy a Scud missile target. Output from the two models was highly similar in terms of overall patterns of workload throughout the mission. In both instances, workload was greatest during the last two minutes of the mission when final decision regarding target presence and location and weapon release needed to be made. Estimates of overall and peak workload from each model were also indistinguishable. The one area in which the models differed was in the component workload estimates obtained for four of the ten functions during the mission. The Micro Saint estimates were consistently somewhat higher than those provided by TAWL, an outcome largely attributable to the differential manner in which the transition periods between tasks are handled by the two models. In sum, the two modeling tools yielded similar results in an overall or gross level, but differed on a fine-grained level, indicating that Micro Saint is much more versatile and flexible than TAWL.

14. SUBJECT TERMS
Operator Workload Measurement, Human Factors, Human Performance Assessment

15. NUMBER OF PAGES
44

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

THIS PAGE INTENTIONALLY LEFT BLANK

PREFACE

This effort was conducted by the Human Interface Technology (AL/CFHP) and the Crew Systems Integration (AL/CFHI) branches of the Armstrong Laboratory at Wright-Patterson Air Force Base, Dayton, Ohio. The project was completed under Work Units 71841425, "Operator Workload Assessment," and 71841044, "Crew-Centered Aiding for Advanced Reconnaissance, Surveillance, and Target Acquisition." Logicon Technical Services, Inc. (LTSI), Dayton, Ohio, provided support under contract F41624-94-D-6000, Delivery Order 0004. Mr. Donald Monk was the Contract Monitor.

The authors wish to acknowledge the support of the Air Force Theater Missile Defense Attack Operations Program Office (ASC/FBXT). In addition, the following individuals should be recognized for their assistance throughout the duration of the project. Gary B. Reid was the task manager, and Gilbert G. Kuperman helped direct the project. Robert Smith of LTSI was responsible for the task analysis of the mission scenario.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

One method for assessing system performance that has witnessed recent widespread growth in popularity is computer task network simulation (Hendy, 1994a). In essence, this technique involves decomposing an activity into individual tasks and simulating their completion via computer so that the impact of proposed modifications on system and operator performance can be evaluated. This type of modeling approach has a number of advantages. First, the effects of proposed modifications on operator performance and workload can be evaluated before the alterations are made; hence, if the model indicates that performance or workload might be adversely affected, potentially disastrous situations can be averted. Second, the computer model can be executed without the expense of constructing a prototype and running experimental tests with human subjects. Third, the computer model can be much more easily modified than a physical model. Inputs to the computer model can easily be altered as additional information (e.g., performance data, task durations, etc.) becomes available. The model can also be readily changed to reflect other proposed modifications to the system. The chief problems with the task network modeling approach, as revealed by a survey distributed to attendees of a workshop on Task Network Simulation for Human-Machine Systems Design held at the Defence Research Agency in Farnborough, U.K. (Hendy, 1993, 1994a), concern the amount of time needed to learn how to use a particular model; inadequate validation of the predictive ability of the models (in terms of both the task timeline and the performance/workload measures); and the poor user interfaces of many computer modeling tools.

Task network simulation has become a particularly widely used technique within the Department of Defense (DoD). In fact, in 1991 the Deputy Secretary of Defense sought to strengthen the application of modeling and simulation in the DoD to promote the effective use of modeling and simulation in training and military operations and in research and development (Kameny, 1995). As part of this initiative, the Defense Modeling and Simulation Office (DMSO) was created in June of 1991 to serve as a center for information concerning DoD modeling and simulation activities. Numerous examples of defense related applications of task simulations testify to the growing recognition of the utility of modeling and simulation to the DoD. Two tools that are frequently used to model crewmember activities and their concomitant performance/workload demands are Task Analysis/Workload (TAWL; Hamilton, Bierbaum, &

Fulford, 1991) and the microcomputer version of the Systems Analysis of Integrated Networks of Tasks (Micro Saint, 1992).

TAWL Modeling Tool

The TAWL methodology was originally developed during the concept exploration and definition phase of the system development process for the Army's Light Helicopter Family (LHX) aircraft to compare the workload of one- and two-crewmember configurations of the LHX. It was specifically equipped to predict operator workload using the techniques developed by McCracken and Aldrich (1984). Their approach to workload is consistent with Wickens' multiple resource theory, which proposes that humans have not just one but several different information processing resources that can be tapped simultaneously in the completion of a task (Wickens, 1984). Under the McCracken-Aldrich approach, workload is viewed as a multidimensional construct that can be divided into five separate components. At any given time, the workload experienced by an operator may stem from one or more of these sources. The five components include visual, auditory, kinesthetic, cognitive, and psychomotor workload. The workload associated with a given task can be estimated by rating each component separately on interval scales developed by Bierbaum, Szabo, and Aldrich (1987) that range from 0 (low workload) to 7 (very high workload). Descriptions of the interval scales corresponding to each of the five workload components can be found in Table 1. For a task, any combination of ratings can result, such that the workload associated with some components might be very high while the workload for other components might be low or nonexistent.

Table 1

Interval Level Workload Component Scales Developed by Bierbaum, Szabo, and Aldrich (1987)

VALUE	DESCRIPTORS
	VISUAL WORKLOAD-UNAIDED
1.0	Visually register/detect
3.7	Visually discriminate
4.0	Visually inspect/check
5.0	Visually locate/align
5.4	Visually track/follow
5.9	Visually read
7.0	Visually scan/search/monitor
	AUDITORY WORKLOAD
1.0	Detect/register sound
2.0	Orient to sound (general)
4.2	Orient to sound (selective)
4.3	Verify auditory feedback
4.9	Interpret semantic content (speech)
6.6	Discriminate sounds
7.0	Interpret sound patterns
	KINESTHETIC WORKLOAD
1.0	Detect discrete activation of a switch
4.0	Detect preset position or status of an object
4.8	Detect discrete adjustment of a switch
5.5	Detect serial movements
6.1	Detect conflict between kinesthetic and visual cues
6.7	Detect continuous adjustment of a switch
7.0	Detect continuous adjustment of controls
	COGNITIVE WORKLOAD
1.0	Automatic (simple association)
1.2	Alternative selection
3.7	Sign/signal recognition
4.6	Evaluation/judgment (consider a single aspect)
5.3	Encoding/decoding, recall
6.8	Evaluation/judgment (consider several aspects)
7.0	Estimation, calculation, conversion
	PSYCHOMOTOR WORKLOAD
1.0	Speech
2.2	Discrete actuation (button, toggle, trigger)
2.6	Continuous adjustive (flight/sensor controls)
4.6	Manipulative
5.8	Discrete adjustive (rotary, vertical thumbwheel, lever position)
6.5	Symbolic production (writing)
7.0	Serial discrete manipulation (keyboard entries)

Prior to executing a model in TAWL, the user must identify a mission of interest and decompose it into progressively smaller units referred to as phases, segments, functions, and tasks. The task represents an event or activity that can be specified in terms of a verb-noun combination (e.g., check gauge, select sensor, set range). It is the fundamental unit of analysis in TAWL. Performance times for each task are estimated as is the workload experienced by the crewmember who completes the task. The model is developed by delineating function decision rules that control the sequencing of tasks within each function as well as segment decision rules that govern the sequencing of functions within segments. Finally, the model is executed using the TAWL Operator Simulation System (TOSS) computer software. The simulation produces estimates of each crewmember's visual, auditory, kinesthetic, cognitive, and psychomotor workload during each half-second period of the mission. When multiple tasks are performed simultaneously, the workload for a particular component is the sum of the ratings across the tasks being completed at that moment in time. Hence, so-called overload conditions with ratings that exceed 7.0 may occur throughout the mission. In this way, the TAWL/TOSS system can be used to identify (1) periods of high workload, (2) crewmembers who experience excessive workload, and (3) components with unusually high workload. This information can subsequently be used to determine the feasibility of adjusting the distribution of tasks throughout the mission, among crewmembers, or among different information processing resources in an attempt to moderate workload levels.

The TAWL/TOSS methodology has been used to study crewmember workload in several investigations. In the initial application, for example, 29 LHX scout and attack mission segments were analyzed (McCracken & Aldrich, 1984). Three different LHX configurations were examined: (1) one crewmember with no automation; (2) one crewmember with automation; and (3) two crewmembers with no automation. A comparison of operator workload in one-crewmember stations with and without automation revealed that the automation considerably reduced the number of occurrences of excessive workload (defined as a component rating greater than 7.0). Overload conditions not only were briefer in duration but also were confined to only three segments with the introduction of automation. A comparison of one- and two-crewmember stations with no automation indicated that the introduction of a second crewmember eliminated excessive workload demands completely in 7 of the 29 segments and reduced them considerably in many other instances (e.g., during functions involving flight control, 193 instances of overload were reduced to 4). Nevertheless, the presence of excessive workload in the 22 remaining

segments suggested that some automation would be required, even in a two-crewmember configuration, to moderate the demands placed on each crewmember.

In another application of TAWL/TOSS, the methodology was used to conduct a task analysis of a UH-60 combat mission (Bierbaum, Szabo, & Aldrich, 1989). Nine phases, 34 segments, 48 functions, and 138 tasks were included in the analysis. The resulting baseline model was used to evaluate the total workload experienced by each crewmember for the current UH-60 aircraft so that the impact of proposed modifications to the aircraft on crewmember workload could later be evaluated. Elements of the model were later incorporated into an investigation designed to assess the predictive validity of computer modeling (Iavecchia, Linton, Bittner, Jr., & Byers, 1989).

In the ensuing validation study, operator workload in a UH-60A Black Hawk simulator was compared to the workload estimates derived from the TAWL/TOSS computer simulation during each segment of the mission. The analysis was conducted by computing and comparing two measures of workload derived from either operator ratings or TAWL output: Overall Workload (OW) and Peak Workload (PW). Following the flight simulation, operators were asked to provide both a rating of the overall amount of workload (OW) and the peak workload (PW) they had experienced during each segment on scales ranging from 0 (very low workload) to 100 (very high workload). In terms of the TAWL/TOSS computer simulation, OW was derived for each half-second interval in the mission by *averaging* across all five component workload estimates; a segment OW measure was then obtained by averaging all of the means within a segment. PW was derived by *summing* the five component workload estimates at each half-second interval and then selecting the maximum or peak workload within the segment. The results revealed that correlations between TAWL-based predictions and crew results were substantial for OW ($\underline{r} = .82$, $\underline{p} < .01$), but somewhat lower for PW ($\underline{r} = .62$, $\underline{p} < .05$). Further, despite the high degree of association, TAWL-based predictions of OW consistently underestimated the ratings provided by human crewmembers.

Micro Saint Modeling Tool

Micro Saint is another modeling tool that has frequently been applied in defense-related assessments. Of the many computer software packages that support task network modeling, it

has proven to be one of the most popular (Hendy, 1994a). The development of Micro Saint began in 1984 when the U.S. Army Medical Research and Development Command sponsored Micro Analysis and Design to develop a user-oriented simulation system that could be run on a microcomputer (Laughery, 1989). What evolved was a general purpose modeling tool targeted primarily for a human engineering audience. While it was not designed for the specific purpose of analyzing operator workload, Micro Saint's versatility makes it perfectly amenable to such analyses. Micro Saint's basic operator interface is a graphical interface which allows information to be input via typing, pointing and clicking with the mouse, or selecting options from available menus. Briefly, a model is constructed in Micro Saint by (1) drawing the tasks on the screen with the tools provided by Micro Saint, (2) entering task attributes such as workload and the mean, the standard deviation, and the shape of the distribution (e.g., normal, gamma, exponential) of the task completion times, and (3) establishing pathways to connect the tasks and control their sequencing. The task attributes are used to depict operator or system performance, whereas the pathways represent the relationships among the tasks in the network. Many different routes through the network become possible as a result of both the user-defined probabilistic and tactical branching between tasks and the variability in task completion times. Hence, each execution of the model will yield different results. Because variability is built into the network, the results of repeated simulations are likely to be indicative of the performance of real-world systems which are themselves characterized by human operator variability.

In a study of human operator workload and cockpit design, Laughery, Drews, Archer, and Kramme (1986) used the Micro Saint modeling tool to simulate four alternative cockpit designs for a future attack helicopter: a generic LHX, a Furness wide model which simulated a wide field-of-view virtual display, a Furness medium which simulated a more limited field of view, and a two-man Apache. Their goal was to assess the effects of the alternative designs on operator workload during anti-armor engagement, a particularly demanding portion of the mission. One technique that the authors used to assess operator overload, in addition to examining component overloads, was to analyze the proportion of time that the operator was unable to update situational awareness outside the cockpit because of excessive visual attention demands. In terms of the computer simulation, the situational awareness task was halted whenever combined visual attention demands exceeded 5.0 on the McCracken-Aldrich scale for visual workload. The results of the simulations revealed that the operator was unable to update situational awareness 60% of the time or more in the generic LHX, but less than 27% of the time

in the other three aircraft. These and other outcomes provided compelling evidence that the generic future attack helicopter would be too demanding of visual attention in comparison to alternative aircraft designs.

In a similar type of application, Ford, Manton, and Hughes (1990) used Micro Saint to assess the workload of seven members of a helicopter and shipboard crew preparing a Royal Australian Navy Seahawk helicopter for an anti-submarine warfare sortie flown from the ship. In particular, because some members could not begin tasks until other members of the team were finished, the intent was to examine the amount of "idle" or unallocated time for each individual during the 150-min mission. Two versions of the model were constructed: one in which the task networks for all seven members began at the same time, and one in which each member's network of tasks did not begin until needed. Each model was executed 100 times. The results of the simulations revealed that the average amount of idle time for the three members of the team who experienced the greatest amount of unallocated time (approximately 50 min or more for each individual) was reduced substantially when they were not called in to perform until absolutely necessary.

Finally, as with the TAWL/TOSS modeling tool, attempts have been made to assess the validity of Micro Saint models. A study conducted by Lawless, Laughery, and Persensky (1995) represents one such endeavor. These authors examined the feasibility and validity of task network modeling to predict the human performance effects of nuclear power plant modifications. Specifically, Micro Saint models were used to examine the difference between the "paper procedures" currently followed in the control room and the new "computerized procedures" that were under consideration but had not yet been implemented. At the same time, traditional experimental tests with human subjects were being conducted in a nuclear power plant control room environment at North Carolina State University to evaluate whether "paper procedures" differed from "computerized procedures." The primary goal of the study was to establish the predictive validity of task network modeling by determining whether the results of the Micro Saint simulations matched those from the experimental tests.

Both paper and computerized procedures for a normal regulatory maneuver and two different accident scenarios were evaluated in both the experimental study and the Micro Saint simulation, providing a total of six conditions in each study. The normal operating conditions

involved a routine change of power operation. The two accident scenarios represented a small break loss of cooling accident (LOCA) and a steam generator tube rupture (SGTR). In all three cases, the dependent variable of interest was the time required by the team to complete the preliminary and final phases of the task. Task performance times for the "paper procedures" Micro Saint model were generated from available empirical data. Comparable times for the proposed "computerized procedures" were developed via expert judgment based on the estimated impact of the new procedures on each of the tasks. Each Micro Saint model was executed 5000 times.

A direct comparison of the "computerized" task performance times from the experimental study and those predicted by the Micro Saint simulation for both the preliminary and final procedures of the three scenarios revealed that the two sets of results were significantly different only in the case of the LOCA accident scenario (both preliminary and final procedures). In both cases, the performance times obtained in the experimental study were longer than the model's predicted times. In the two remaining scenarios, the average performance times predicted by the Micro Saint model did not differ from those actually obtained in the empirical study. Thus, the model values matched the empirical values in four of the six possible conditions. The authors concluded that while task network models are easily constructed and readily modified, their predictive validity is not yet sufficiently high to permit a definitive declaration of the success of the modeling approach.

The Present Study

The present study was an attempt to use the modeling approach to examine the workload experienced during an F-15E target acquisition ("Scud hunt") mission by the weapons system operator (WSO), the crewmember occupying the back seat of the aircraft. There were three primary goals of the current research. First, the chief aim was to develop both Micro Saint and TAWL/TOSS computer models depicting the mental workload associated with the WSO's tasks during the mission. A second goal was to compare the output of the two computer task simulation models. Finally, the output from the model that is selected will be correlated with the workload derived from simulated missions (laboratory simulations and military field exercises) to determine the predictive validity of the technique. The current report covers the first two of these goals. The results of the third purpose will be documented at a later date. As in other

studies of this nature, if the modeling technique proves valid, it will be used to study and predict the effects of various modifications within the scenario (e.g., changes in image resolution, task allocation, etc.) on operator workload, ultimately in lieu of experiments with human subjects.

METHOD

Scenario Development

The mission used in the present study was designed to portray the tasks that a WSO must complete during a Scud hunt mission. In general, the WSO is primarily responsible for studying the available radar imagery in order to detect, locate, and designate the target. When the 5-6 min scenario begins, the F-15E has already been diverted to investigate a potential Scud missile target located 30-40 nautical miles (nmi) away. For analytical purposes, the mission was divided into three broad segments: target detection, target destruction, and damage assessment. These three segments were further subdivided into ten functions, each of which consists of a set of tasks necessary for the completion of the designated activity:

(1) Initialize air-to-ground (A/G) mode--the pilot prepares the aircraft's displays for air-to-ground as opposed to air-to-air delivery mode

(2) Perform inertial navigation system (INS) update--the WSO completes a set of tasks to ensure accurate target positioning

(3) Obtain patch map for orientation--the WSO obtains a radar image of the area in the vicinity of the suspected target at a resolution suitable for overview of the scene (8.5 ft resolution in this scenario)

(4) Obtain patch map for detection--the WSO obtains a second image at a finer resolution (4 ft x 6 ft in this scenario) that will permit target detection and subsequent designation

(5) Verify weapon status--the WSO verifies that weapons are available and ready for use

(6) Detect target--the WSO makes a final determination of the presence/absence and location of a target in the scene

(7) Designate target--the WSO inputs target location data by positioning a cursor over the target in the image

(8) Track and identify target--the WSO views Forward-Looking Infrared (FLIR) imagery to identify the target and designate its location with greater accuracy if necessary

(9) Release weapon--the pilot releases the weapon

(10) Assess damage--the WSO inspects the FLIR imagery to view and assess the extent of the damage

Finally, each function was further subdivided into one or more individual tasks necessary for its completion. Because the focus in the present study was on the mental workload experienced by the WSO rather than the pilot, any pilot tasks that had to be finished before the WSO could begin certain tasks were included in the model only as timeholders.

The particular mission that was used was designed to suffice not only for model construction in the present study but also for simulations with human subjects to be conducted in the F-15E simulator located in the Crew-Aiding and Information Warfare Analysis Laboratory (CIWAL) at Wright-Patterson Air Force Base in Dayton, OH. Ultimately, the output from the present computer simulations will be compared with the results of these laboratory simulations as a first step toward determining their predictive validity. In addition, because it was intended to serve as a baseline model from which more complex models could be constructed in the future, the mission was somewhat simplified in nature. For example, it was assumed that the target was always present at the coordinates that the pilot had received and that no other threats (e.g., surface-to-air missiles, anti-aircraft artillery) were present.

In general, the equipment required by the WSO for completing each task consists of four multi-purpose displays (MPDs) and two hand controllers (HCs). The MPDs are two color and two monochromatic computer monitors arranged in a row from left to right in the WSO's station. Each MPD is surrounded by 20 push buttons (PBs). The majority of the WSO's tasks can be accomplished by pressing the correct PB on the appropriate display. Other tasks may be

completed via the left and right HCs, each of which contains eight switches/buttons/triggers. The left HC is used for tasks accomplished with any display on the two leftmost MPDs, whereas the right HC is used for the two rightmost MPDs. In most instances, a task can be completed via either the PBs or the HCs, depending upon the WSO's preference.

Once the tasks required for the completion of the scenario were identified, descriptions of each were written to facilitate subsequent derivation of workload and task duration estimates. These descriptions were written with the aid of several F-15E manuals, including the Flight Manual and the Nonnuclear Weapon Delivery Manual.

Workload Estimation

The workload associated with each task was estimated via the McCracken-Aldrich (1984) approach using the interval scales later developed by Bierbaum, Szabo, and Aldrich (1987). The descriptions of the interval level scales for each workload component were used in conjunction with the task descriptions that had been written in order to estimate the workload associated with each task. Because auditory tasks (e.g., communications to, from, and within the cockpit) were not included in this initial simplified model, the auditory workload component for all tasks was designated as 0. Many of the tasks involved simple button/switch activations from either the MPDs or the HCs; hence, a majority of the tasks received psychomotor workload ratings of 2.2 and kinesthetic ratings of 1.0. All component workload ratings were estimated twice (separated by approximately three months) in order to permit an assessment of the reliability of the workload estimation technique.

The component workload ratings were used not only to identify component overloads but also to derive two other indices of workload described earlier: Overall Workload (OW) and Peak Workload (PW). In the present study, these measures were derived by first computing averages and sums of the component workload estimates at each half-second interval of the mission. Subsequently, each function included in the mission was then examined to determine the OW and PW associated with that function.

Task Durations

Means and standard deviations for the duration of each task were derived via three techniques: (1) examination of the task durations used in other similar task analyses (e.g., Hendy, 1994b; McCracken & Aldrich, 1984); (2) examination of empirical data from laboratory investigations of both full- and part-task missions (e.g., Kuperman, Wilson, & Davis, 1993; Kuperman, Wilson, & Perez, 1988); and (3) expert judgment. In particular, the task completion time of 400 milliseconds (ms) for operating a push button or toggle, an action which figured into nearly all of the tasks included in the current study, was taken from a table of task completion times provided by Hendy (1994b). The report by McCracken and Aldrich (1984) was helpful in determining task duration times associated with such tasks as cursor positioning and damage assessment. The results of laboratory investigations were used primarily to determine the task completion times for image evaluation and detection. When estimates could not be derived by either of these two techniques, expert judgment was used. As in other studies (e.g., McCracken & Aldrich, 1984), a half-second transition time was added to the end of each task. Finally, because information regarding the standard deviations in task duration times was altogether unavailable, all estimates of standard deviations were derived from expert judgment.

Micro Saint Simulation

The F-15E scenario was run first with the Micro Saint modeling tool for Windows (Release 1.3, Build R). One feature unique to Micro Saint is the ability to designate not only task duration means and standard deviations but also the type of distribution from which the task completion times are sampled during each simulation run. In the current model, the gamma distribution was used for all tasks involving discrete activations (e.g., PB and HC actions). This type of distribution is ideal for tasks such as discrete activations that generally cannot be performed much more quickly than the mean but could potentially take much longer. The normal distribution was used for all other types of tasks, which could conceivably be completed either more slowly or more quickly than average (e.g., positioning the cursor, studying the map, verifying weapon availability). A second unique feature of Micro Saint is the flexibility that it allows in the designation of workload estimates. Because each workload component is defined and entered as a separate variable, the number of workload components is virtually unlimited.

Hence, this feature enabled the psychomotor workload component to be further subdivided into right hand and left hand psychomotor components in Micro Saint.

After the entire model had been entered in Micro Saint, the scenario was run 100 times, the number of iterations generally required to obtain stable results (Siegal & Wolf, 1969). Component workload estimates were obtained for each half-second interval of each mission. The data file was edited and transported to a PC-based version of the Statistical Analysis System (SAS, 1992) for further analysis.

TAWL Simulation

Following the completion of the Micro Saint simulation, the F-15E mission was adapted for use with the TAWL modeling tool. First, because TAWL does not require either the standard deviations or the type of distribution associated with task performance times, these were dropped from the TAWL version. Second, the task completion times themselves were revised since TAWL requires that all inputs be multiples of .5 sec (e.g., the 400 ms "button press" time had to be changed to 500 ms or .5 sec). In addition, the task completion times had to be revised further since TAWL automatically attaches a half-second transition time to each task, whereas it must be added manually in Micro Saint. Third, the subsystem(s) pertaining to each task had to be identified for entry into TAWL. Briefly, the subsystem identifies the equipment associated with the performance of a task (e.g., radar, navigation, weapons, etc.). It can be defined as narrowly or broadly as needed. This entry is used during the model simulation to determine the number of subsystem overloads that take place. Fourth, tasks had to be classified as either discrete/continuous and fixed/random. Continuous tasks are those tasks whose magnitude of performance determines the magnitude of the ensuing system response (e.g., pushing forward on the stick to control pitch). The intensity of performance of discrete tasks does not affect the magnitude of the resulting system response (e.g., pressing a button to select the type of sensor). With respect to the fixed/random dimension, fixed tasks must be performed at a fixed time in relation to the performance of other tasks (e.g., the cursor must be positioned before the target can be designated). The time of occurrence of random tasks cannot be determined a priori. Such tasks may occur at any time during a function, depending on factors such as crewmember preference and current workload (e.g., a flight gauge may be checked whenever and as often as time permits).

Finally, the decision rules for progressing from one task and function to another had to be altered for input in TAWL. Because TAWL does not permit probabilistic and tactical branching as Micro Saint does, a single representative scenario was developed specifically for TAWL. This model was input and run once. (When there are no random tasks in the model, the output will not vary). Component workload estimates were obtained for each half-second interval of each mission. The data file was edited and transported to SAS for further analysis.

RESULTS

Reliability of Workload Estimates

Prior to comparing the output from the two computer models, the reliability of the workload estimation technique itself was assessed. The workload associated with a total of 6 pilot and 45 WSO tasks was included in this analysis. Although the concern in the present study was to determine the mental workload experienced by the WSO, the workload of 6 pilot tasks that had to be completed before the WSO could begin certain tasks was also estimated. Because the intent was to assess the reliability of the workload estimation technique per se, these pilot workload estimates were included in the reliability assessment. However, although these estimates might potentially figure into future models, they did not enter directly into any of the models constructed in the present study.

Correlations.

As a preliminary method for exploring the general relationship between the workload estimates over time, the correlations between the two sets of ratings were examined. Separate correlations were computed for each of the five workload components: visual ($r = .89$), auditory ($r = 1.00$), kinesthetic ($r = .75$), cognitive ($r = .80$), and psychomotor ($r = .44$). All of the correlations were statistically significant at $p < .0001$, except for the correlation between psychomotor ratings ($p < .001$). The nature of the correlations indicates that the workload estimates for the visual, auditory, kinesthetic, and cognitive components were highly consistent over time; whereas the estimates for the psychomotor component were considerably more discrepant.

Intraclass correlation coefficients.

The primary method for determining the reliability of the workload estimation technique was to examine intraclass correlation coefficients (ICC) for each workload component. Product-moment correlation coefficients are useful for ascertaining whether there is an *association* between the two sets of ratings, but they fail to indicate whether the two ratings for a given task actually *agree* (i.e., the extent to which the two ratings are identical; Bartko & Carpenter, 1976; Jones, Johnson, Butler, & Main, 1983). The ICC is a measure of reliability that does estimate the extent of agreement (Bartko & Carpenter, 1976). In essence, the ICC is a comparison of the variances within (MS_W) and between (MS_B) ratings that is derived by conducting an analysis of

variance. When the number of ratings per entity is two and there are no missing ratings, ICC = $(MS_B - MS_W)/(MS_B + MS_W)$ and ranges from -1.00 to 1.00, with larger values representing greater agreement. If the variability within ratings (i.e., the ratings for each task across time) is small relative to the variability among ratings (i.e., the workload ratings across the 51 tasks), the ICC will be large. If the two sets of ratings agree perfectly, the variance within ratings will be 0, yielding an ICC of 1.00 (provided the variability between ratings is non-zero). The $\underline{F}$ test from the analysis of variance further indicates whether the ICC is statistically different from 0, a value representing no agreement.

The intraclass correlation coefficients and their associated $\underline{F}$ tests for each workload component can be found in Table 2. Because no auditory tasks were included in this simplified mission scenario, the auditory workload component was omitted from the table. The ICC indices for the four remaining workload components were significantly greater than 0. As with the Pearson product-moment correlation coefficients reported earlier, the figures in Table 2 indicate that the visual, kinesthetic, and cognitive workload components were estimated with considerable consistency over time; whereas the psychomotor workload estimates were less reliable. Inspection of the raw ratings revealed that the discrepancy for the psychomotor workload component was chiefly due to the disparate ratings of three tasks (setting range, setting azimuth, and positioning the cursor), each of which is completed at three separate times during the mission. These and other discrepancies in the workload ratings were resolved prior to running the scenario in either Micro Saint or TAWL.

Table 2

Intraclass Correlation Coefficients and F Tests of Significance (df = 50,51) for Four Components of Workload

Workload Component	ICC	$\underline{F}$	$\underline{p}$
Visual	0.89	16.85	.0001
Kinesthetic	0.71	5.98	.0001
Cognitive	0.80	8.91	.0001
Psychomotor	0.43	2.50	.0007

Micro Saint Simulation

Usability.

The first method for assessing each modeling tool involved evaluating its "user friendliness." With the Micro Saint modeling tool, entry of the basic model was easily accomplished in about a half hour. The process is highly intuitive in that each task is drawn using the "task" tool. Task elements are entered by double-clicking on the task icon and typing them in the appropriate slots or selecting from available menu options. Paths between tasks are established by using the "path" tool. The manual that accompanies the software is clearly written and provides a number of different examples of task networks; in addition, extensive on-line help is available.

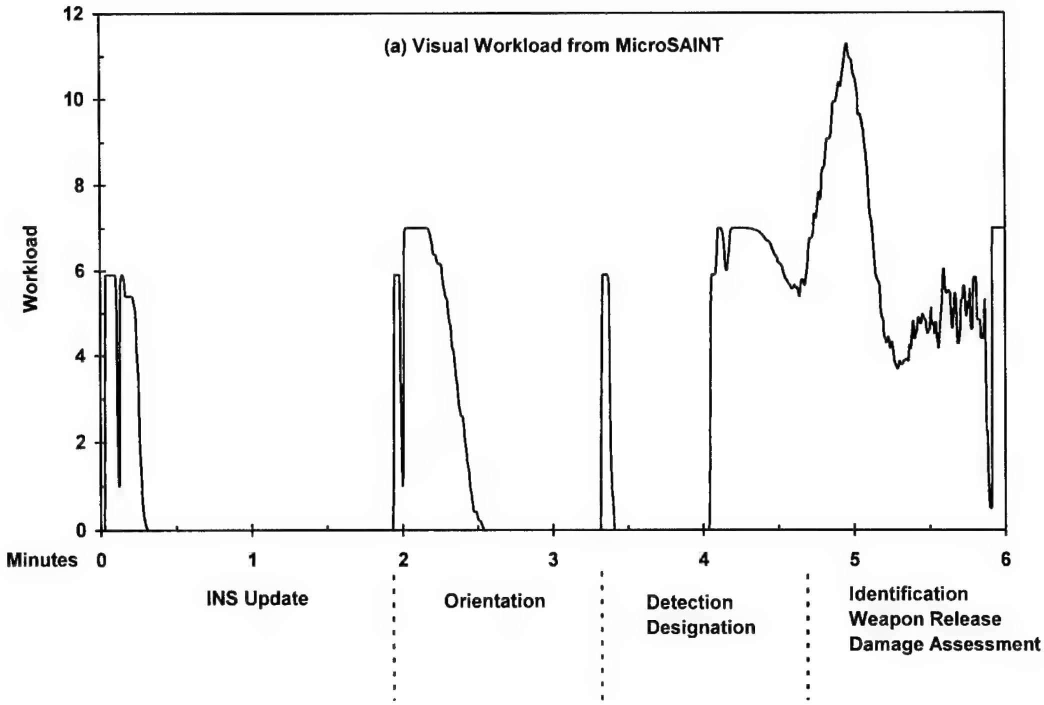
Micro Saint can perhaps best be characterized as a highly flexible modeling tool. Part of its flexibility stems from its generality. For example, many of the inputs are user-defined and are therefore potentially limitless in their capabilities. Any number of workload components can be entered since each component is simply a variable that the user defines. By the same token, Micro Saint is equipped to handle any number of different operators, tasks, variables, and data collection "snapshots." Further evidence of Micro Saint's flexibility is the ease with which tasks and paths can be added and deleted. Micro Saint also enables the user to draw a true task "network": several tasks can occur simultaneously, and a single task might be followed by one of three different tasks depending on tactical or probabilistic user-defined rules (e.g., Task A might be followed by Task B 25% of the time, Task C 40% of the time, and Task D 35% of the time). Several run options are available, including the ability to set all standard deviations to zero (useful for debugging). In addition, the user can enter the exact number of times the model should be run. All of the output from that set of runs is contained within a single file, which provides ease of editing and transport to other statistical and graphics packages.

Although Micro Saint has the capability to perform some simple statistical analyses and graphical displays of the output, this feature is limited. Ultimately, it is easier to transport the raw data to other packages specifically equipped to perform statistical/graphical analyses (e.g., SAS, Microsoft Excel).

Workload analysis.

The output from the Micro Saint simulation was first used to determine that the mission length of the 100 simulated runs ranged from 5.32 to 6.07 minutes (M = 5.58 min, SD = 0.14), a duration that was well within the intended range. Next, the workload estimates from each of the 100 mission runs were used to determine the mean workload during each half-second interval for the "average" mission. The mean component workload estimates are plotted as a function of time in Figures 1a-f (note that the plot of auditory workload was omitted since no auditory tasks were included in this scenario). The plots in Figure 1 are most useful for visually identifying those periods of the mission that are most demanding, as indicated by peak workload estimates. It can be seen that the greatest demands on the WSO's visual and cognitive skills occurred during the last portion of the mission when final decisions regarding target presence and weapon release must be made. Specifically, the peak visual and cognitive workload estimates are observed when target identification, weapon release, and damage assessment are required. The plots in Figure 1 also indicate that whereas the left hand is occupied throughout the mission, the right hand is needed only during the last third of the mission. Hence, most of the psychomotor workload stems from activities requiring use of the left hand. Finally, the gaps in each figure represent "idle" time periods during which the WSO has finished the tasks necessary to successfully fulfill function requirements, but the pilot must continue to complete other flight tasks.

The estimates of visual, auditory, kinesthetic, cognitive, and psychomotor (general) workload portrayed in Figure 1 were used to compute estimates of overall (OW) and peak (PW) workload associated with each of the 10 functions during the mission. These estimates can be found in Table 3. The values of OW indicate that the average of the five workload components was greatest during target designation and target tracking/identification. Consistent with the plots in Figure 1, the sum of the five workload components reached a peak during the last third of the mission when the target was being tracked and identified. Peaks also occurred during the INS update and during target designation.



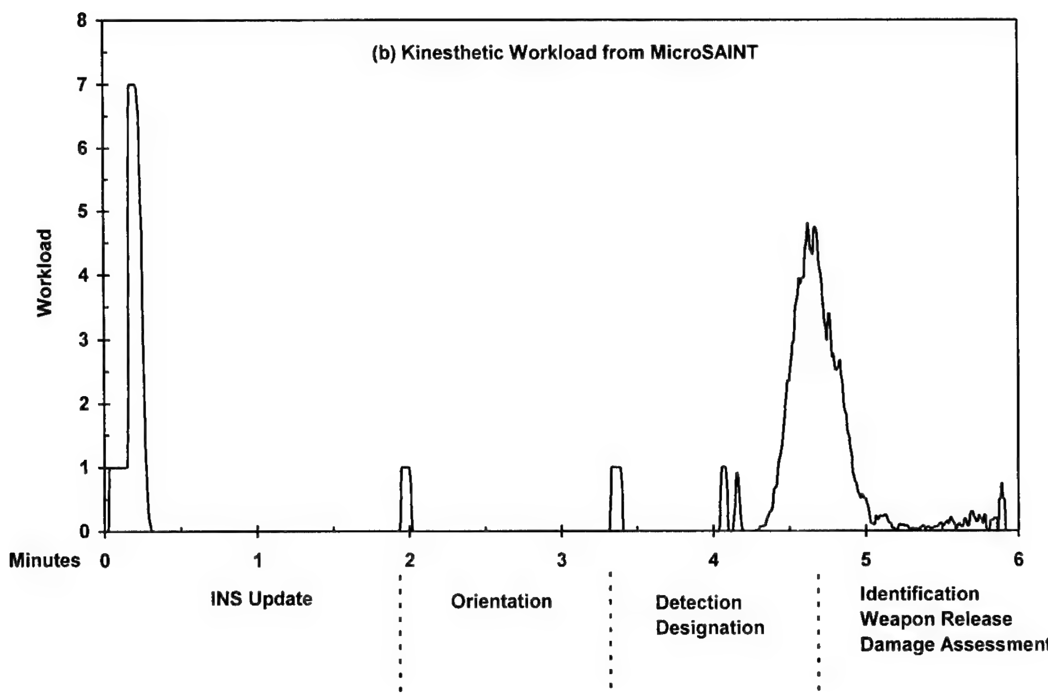
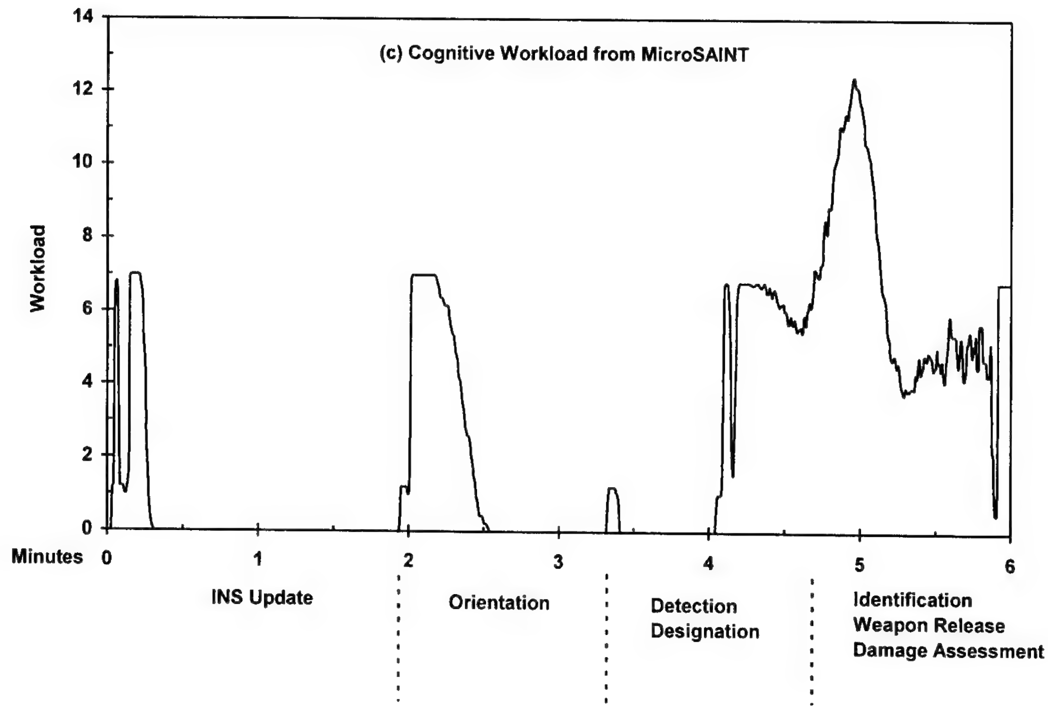
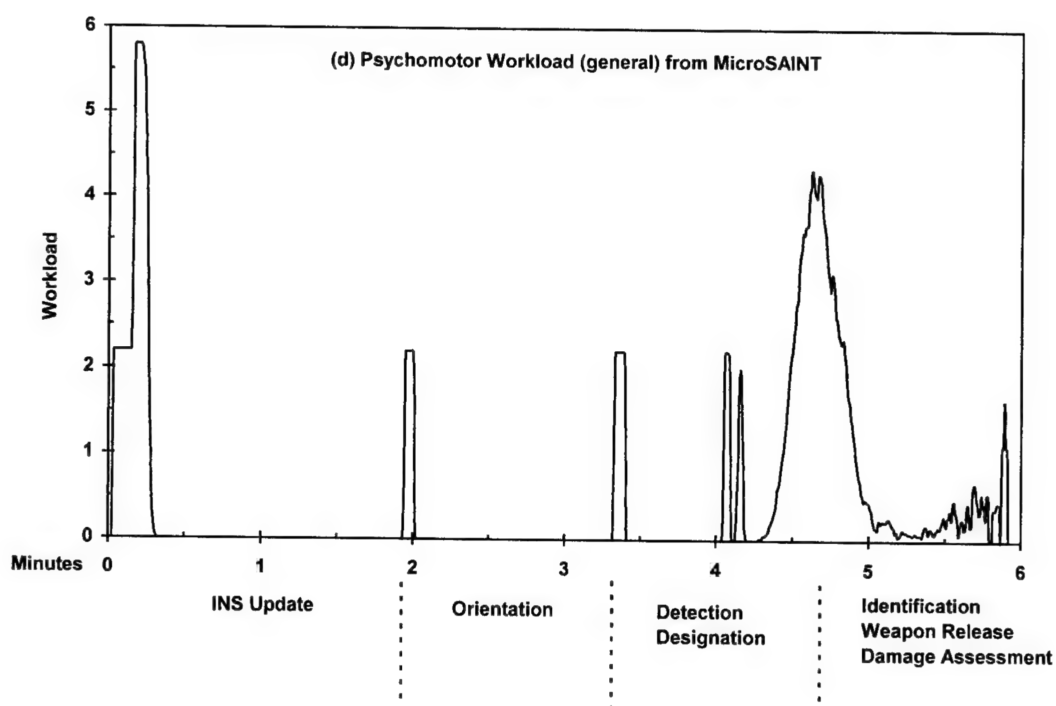
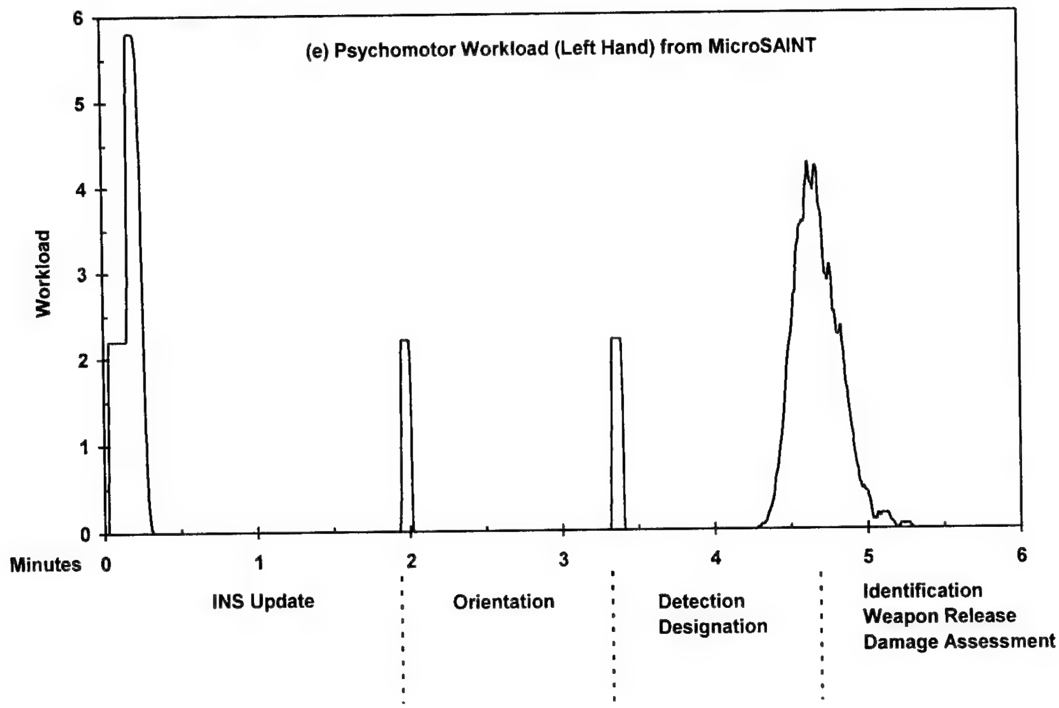


Figure 1. Mean estimates from Micro Saint of (a) Visual, (b) Kinesthetic, (c) Cognitive, (d) Psychomotor (general), (e) Psychomotor (left hand), and (f) Psychomotor (right hand) workload as a function of time. (figure continues)

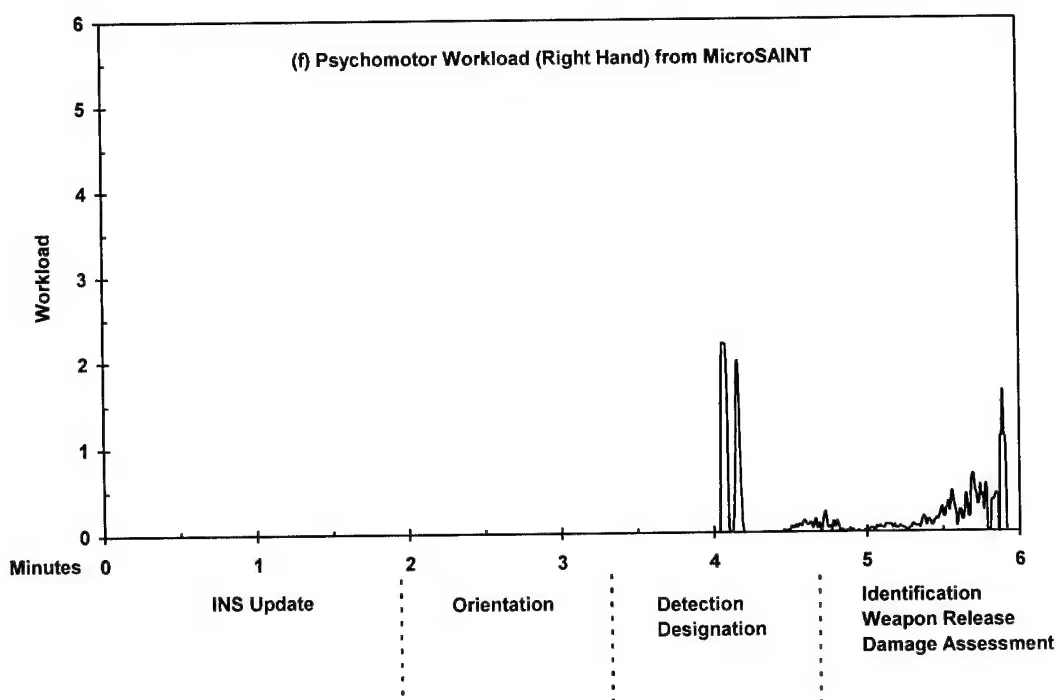




(figure continues)



(e) Psychomotor Workload (Left Hand) from MicroSAINT
Workload
6
5
4
3
2
1
0
Minutes 0
1
2
3
4
5
6
INS Update
Orientation
Detection
Designation
Identification
Weapon Release
Damage Assessment



(f) Psychomotor Workload (Right Hand) from MicroSAINT
Workload
6
5
4
3
2
1
0
Minutes 0
1
2
3
4
5
6
INS Update
Orientation
Detection
Designation
Identification
Weapon Release
Damage Assessment

Table 3

OW and PW Estimates from Micro Saint for each Function in the Scenario

Function	OW	PW
(1) Initialize A/G mode	0.00	0.00
(2) Perform INS update	0.41	25.20
(3) Obtain patch map (orientation)	0.79	14.00
(4) Obtain patch map (detection)	0.20	10.30
(5) Verify weapon status	2.26	13.80
(6) Detect target	2.76	13.80
(7) Designate target	4.20	25.20
(8) Track and identify target	3.98	26.00
(9) Release weapon	0.00	0.00
(10) Assess damage	2.53	13.80

TAWL Simulation

Usability.

As with Micro Saint, entry of the basic model in TAWL required only about a half hour. However, it took nearly eight hours to accomplish a successful TAWL run. The chief reason for this difficulty involved the specification of task performance times in TAWL. These estimates must be entered as multiples of .5 sec, a constraint which is never specifically referenced in the manual that accompanies the TAWL software. Initially, the task performance times that had been expressed in milliseconds in Micro Saint were simply converted into seconds for TAWL (e.g., 900 ms was converted to .9 sec). Subsequently, it was discovered that the model will not proceed properly from task to task as it should unless task performance times are multiples of .5 sec. Needless to say, this was realized only after many frustrating attempts to identify the source of the errors in the output.

There are several other features of TAWL that make it a less than desirable modeling tool. First, the worksheets that must be completed before entering the model in the computer

with the TOSS software are somewhat redundant. The function summary and segment summary worksheets seem to be unnecessary precursors to the function decision rules and segment decision rules worksheets. Second, TAWL requires precise estimates not only of each task duration but also of each function duration. Determining the function duration is cumbersome and redundant, and errors here adversely affect the output. If the function duration that is entered is too short, tasks are omitted from the output. If the function duration is too long, gaps of zero workload appear in the output. A related problem is the fact that TAWL automatically inserts a half-second transition time between tasks, yet this is not referenced in the manual. Third, the transitions within and between input screens in TAWL are not intuitive. The manual must be consulted constantly. The user never knows whether to press "esc," "enter," or "tab." Further, inadvertently pressing the "esc" key at the main menu closes the system entirely. Fourth, when a model is executed in TAWL, each segment of the mission must be run separately. In the output, the clock begins again at 0 for each segment. If the user is interested in examining the mission as a whole, the timeline must be re-computed.

Additional problems might be referred to as a lack of sufficient flexibility. For example, because standard deviations are not required for task performance times, the output will be identical every time the model is executed (as in the current study) unless random tasks are included in the model. This is a serious drawback since proper statistical analyses cannot be conducted in the absence of variability in output from run to run. A desire to run a model a number of times would also be hindered by the fact that TAWL does not have a feature for specifying the number of times the model should be executed. Hence, separate runs would need to be executed, and the output files would have to be manually collated (e.g., with a word processor) prior to analysis. Another form of inflexibility in TAWL arises from what at first appears to be a flexible feature. TAWL enables the designation of workload component specifiers to permit further subdivision of a workload component. For example, the psychomotor component might be divided into right vs. left hand while the visual component might be divided into head-up vs. head-down. During model simulation, TAWL will identify instances where component specifiers clash (e.g., two tasks that simultaneously require use of the left hand). The disadvantage is that TAWL does not also provide estimates of workload separately for each component specifier. Hence, as just one example, the user cannot determine whether the left hand is required more often than the right. Finally, TAWL is limited to a maximum of six workload components and four crewmembers.

Several positive features include the ability to enter an alternative workload equation (e.g., an equation to convert TAWL component ratings to common subjective workload scores) and the ability to designate different overload thresholds. Further, several different types of output are available, including screen output and a numerical data file suitable for export to other statistical/graphics packages. In addition, various reports of function/segment names and decision rules can be generated to facilitate review of the model's accuracy.

Workload analysis.

The output from the TAWL simulation indicated that the mission duration was 5.7 min. Component workload estimates were obtained for each half-second interval of the mission and are plotted as a function of time in Figures 2a-d. The plots in Figure 2 reveal that the peak visual and cognitive workload occurred during target identification, weapon release, and damage assessment. Further, while kinesthetic and psychomotor workload reached a maximum at this time as well, both also exhibited a comparable peak during the INS update portion of the mission. As with the plots from Micro Saint, the gaps in each plot in Figure 2 represent "idle" time during which the WSO has fulfilled all requirements but the pilot has not.

Estimates of OW and PW during each function of the mission, computed from the estimates of visual, auditory, kinesthetic, cognitive, and psychomotor (general) workload obtained from TAWL, can be found in Table 4. The values of OW indicate that the average of the five workload components was greatest during target tracking/identification and target designation. Consistent with the plots in Figure 2, the sum of the five workload components reached a peak during the last third of the mission when the target was being tracked and identified. Peaks also occurred during the INS update and during target designation.

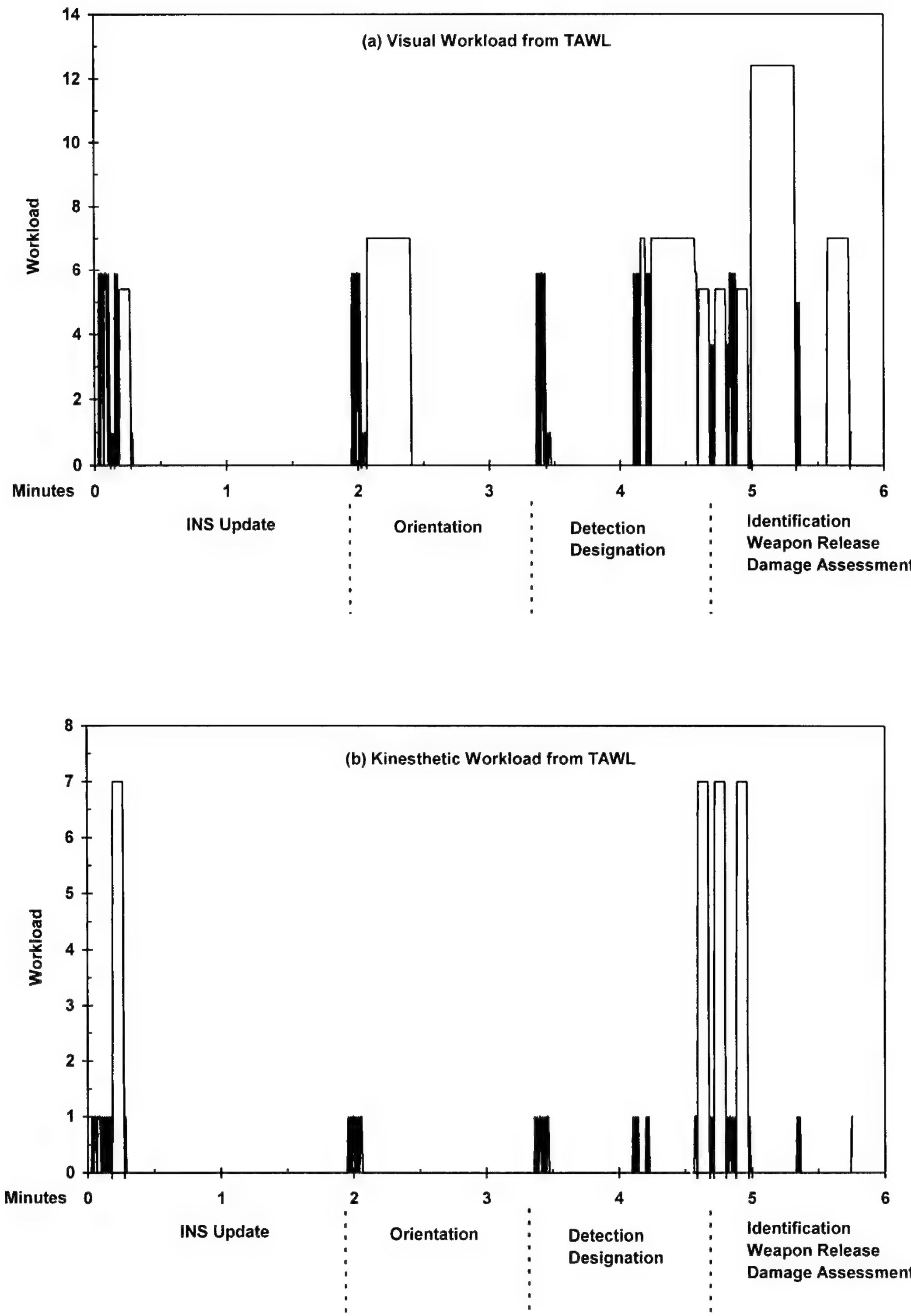
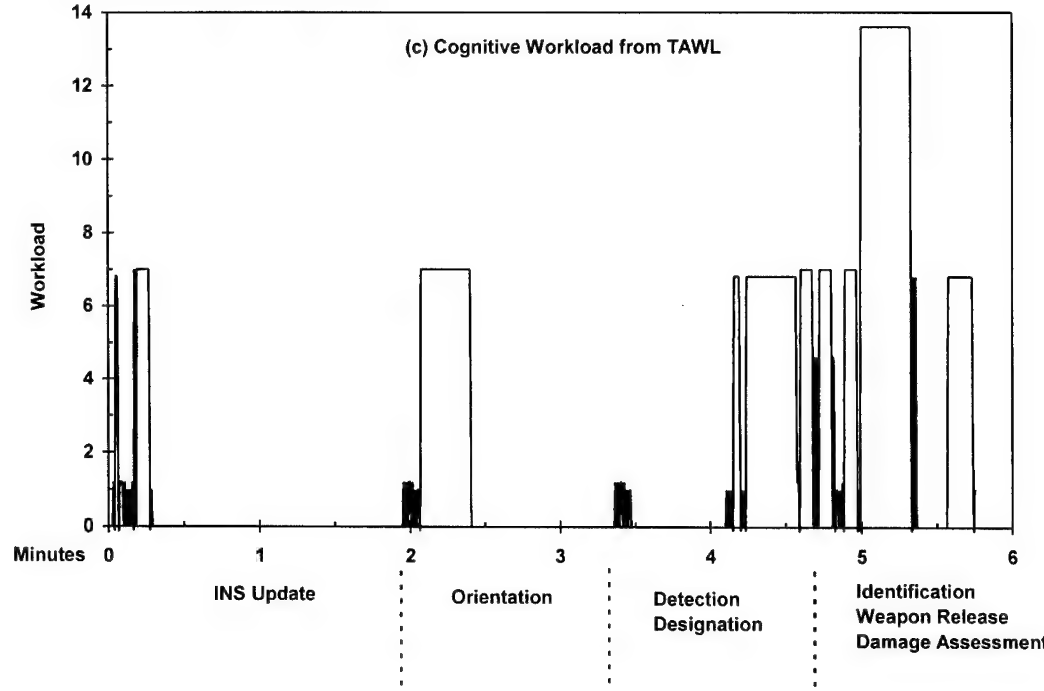
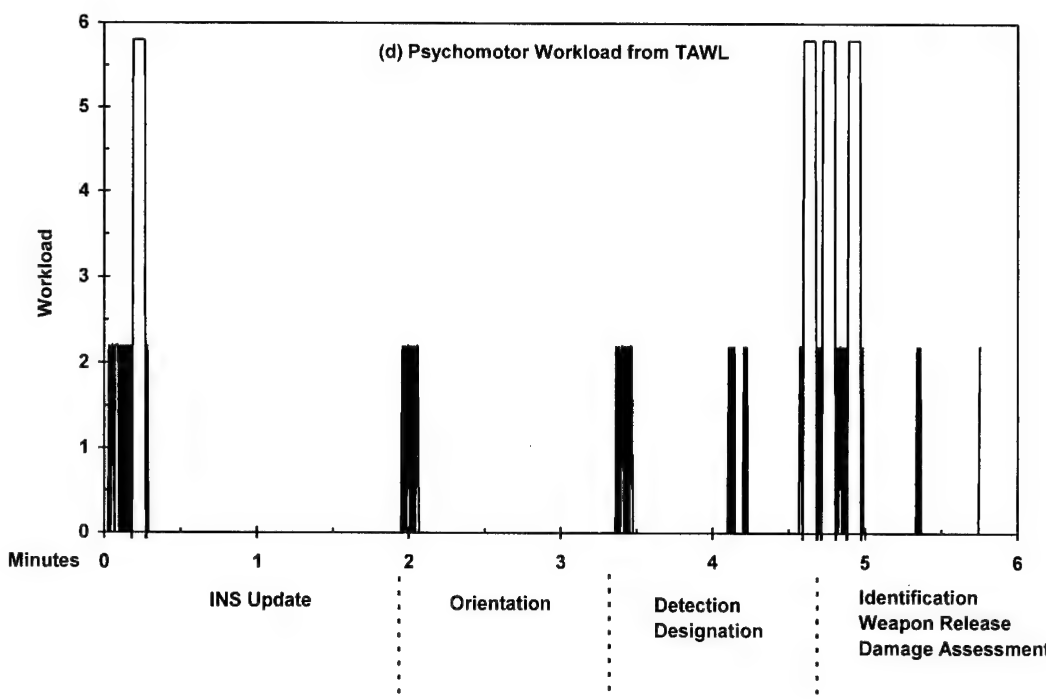


Figure 2. Mean estimates from TAWL of (a) Visual, (b) Kinesthetic, (c) Cognitive, and (d) Psychomotor (general) workload as a function of time. (figure continues)



(c) Cognitive Workload from TAWL
Workload
14
12
10
8
6
4
2
0
Minutes 0 1 2 3 4 5 6
INS Update
Orientation
Detection
Designation
Identification
Weapon Release
Damage Assessment



(d) Psychomotor Workload from TAWL
Workload
6
5
4
3
2
1
0
Minutes 0 1 2 3 4 5 6
INS Update
Orientation
Detection
Designation
Identification
Weapon Release
Damage Assessment

Table 4

OW and PW Estimates from TAWL for each Function in the Scenario

Function	OW	PW
(1) Initialize A/G mode	0.00	0.00
(2) Perform INS update	0.33	25.20
(3) Obtain patch map (orientation)	0.74	14.00
(4) Obtain patch map (detection)	0.15	10.30
(5) Verify weapon status	1.49	13.80
(6) Detect target	2.76	13.80
(7) Designate target	3.61	25.20
(8) Track and identify target	4.21	26.00
(9) Release weapon	0.00	0.00
(10) Assess damage	2.49	13.80

Comparison of TAWL and Micro Saint

OW and PW.

For ease of comparison, the OW and PW estimates from Micro Saint and TAWL that were portrayed in Tables 3 and 4 have been plotted for eight functions from the mission in Figures 3 and 4. The two functions in which all of the tasks were pilot tasks (Initialize A/G mode and Release weapon) were omitted from the figures since the WSO's workload was zero during those time periods. The overall and peak workload estimates from Micro Saint and TAWL were highly similar. In fact, in the case of the peak workload estimates, Micro Saint and TAWL yielded identical estimates. In general, the overall workload estimates from Micro Saint tended to exceed those from TAWL for all functions except target tracking/identification, where the estimate from TAWL was higher, and target detection, where the estimates were identical.

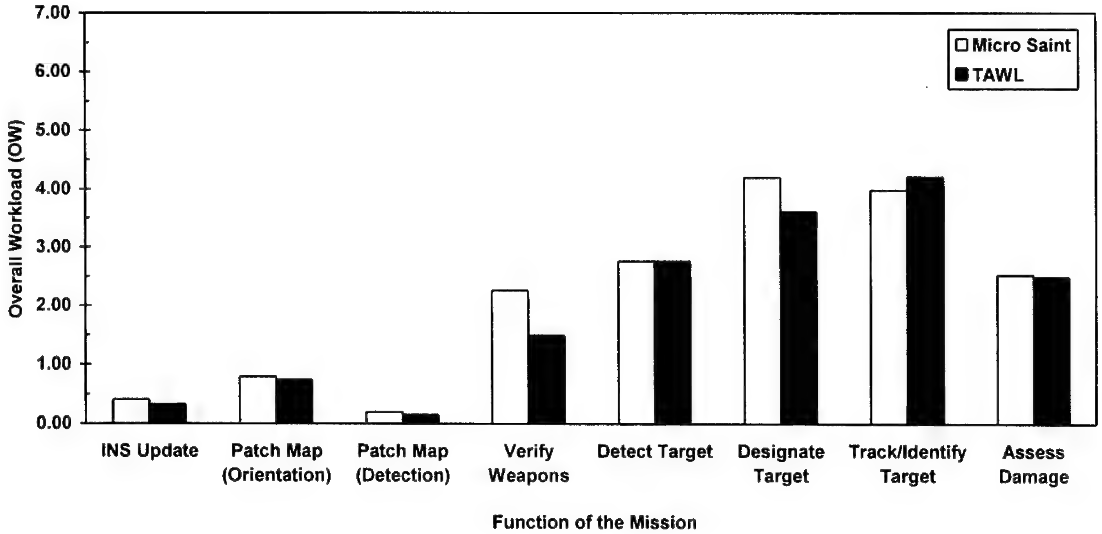


Figure 3. Overall workload estimates from Micro Saint and TAWL for each function from the mission.

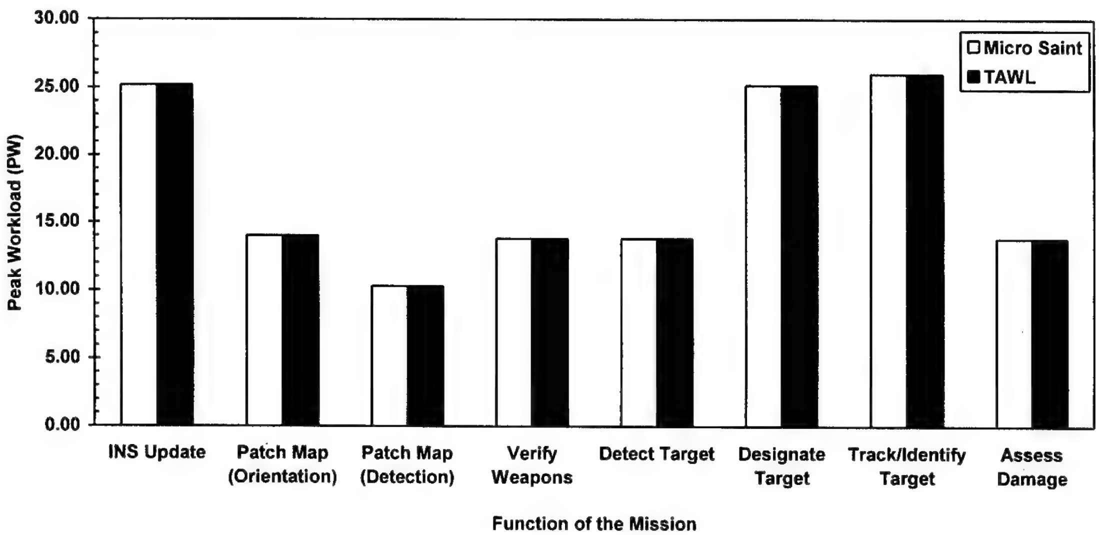


Figure 4. Peak workload estimates from Micro Saint and TAWL for each function from the mission.

The apparent similarity in the OW and PW estimates produced by Micro Saint and TAWL was verified by the results of correlational analyses and an analysis of variance. First, Pearson product-moment correlations between Micro Saint's and TAWL's estimates were $\underline{r} = .98$

(OW) and $\underline{r} = 1.00$ (PW), both of which were statistically significant at $\underline{p} < .0001$. Second, in order to assess further whether the OW and PW estimates provided by Micro Saint and TAWL were statistically different, a 2 (Workload Measure) x 2 (Modeling Tool) analysis of variance was done on the scores that were obtained for the ten functions of the mission. Only the main effect for type of Workload Measure (OW vs. PW) was statistically significant, $\underline{F}\ (1,9) = 22.19$, $\underline{p} < .0011$. Neither the main effect for Modeling Tool nor the interaction between Workload Measure and Modeling Tool was large enough to attain statistical significance ($\underline{p} > .20$). Hence, in terms of estimating OW and PW during a mission, Micro Saint and TAWL can be considered equivalent.

<u>Component workload estimates.</u>

In addition to assessing the similarity of Micro Saint and TAWL with respect to estimates of OW and PW, the ratings on each workload component produced by the two modeling tools were also compared. A separate 2 (Simulation Model) x 5 (Workload Component) analysis of variance was done on the component workload scores for each function. Two functions (Initialize A/G Mode and Release Weapon) were omitted from these analyses since the workload for the WSO was zero while the pilot completed these tasks. In addition, the target detection function was omitted since it consisted of only a single task and therefore provided only one estimate for each workload component (hence, no variability was present to permit the completion of an analysis of variance for this function). The results of the seven Anovas that were conducted are depicted in Table 5. If the output for each workload component from the two models is equivalent, neither the main effect for Simulation Model nor the interaction between Simulation Model and Workload Component should be significant. A main effect for Workload Component would indicate only that the workload ratings differed across the five components. The figures in Table 5 reveal that the output from Micro Saint and TAWL did not differ for three of the functions but did differ significantly for the remaining four: Verify Weapons, Designate Target, Track/Identify Target, and Assess Damage.

Table 5

Results of Seven 2 (Simulation Model) x 2 (Workload Component) Analyses of Variance of Component Workload Ratings

	Effect					
	Simulation Model (A)		Workload Component (B)		A x B Interaction	
Function	F	p	F	p	F	p
INS Update	3.02	NS	19.23	.0001	2.14	NS
Patch Map (Orientation)	.74	NS	68.72	.0001	.52	NS
Patch Map (Detection)	.97	NS	8.43	.0039	1.23	NS
Verify Weapons	15.56	.0010	36.98	.0001	10.41	.0005
Designate Target	224.06	.0001	458.63	.0001	173.33	.0001
Track/Identify Target	120.58	.0001	586.52	.0001	55.45	.0001
Assess Damage	302.97	.0001	437.24	.0001	311.73	.0001

In order to determine which specific components differed between the two models for these four functions, follow-up tests were completed. Specifically, for each function, four separate repeated measures Anovas were conducted to ascertain whether the (1) visual, (2) kinesthetic, (3) cognitive, and (4) psychomotor workload ratings differed according to Simulation Model. The results of these tests are tabulated in Table 6. As can be seen in the table, all of the component ratings produced by Micro Saint and TAWL differed significantly for all four of the functions, with one exception: the cognitive workload ratings produced by the two models did not differ for the weapon verification function.

Descriptive statistics for these functions can be found in Table 7. As can be seen in the table, the mean ratings for each workload component were consistently higher for Micro Saint as opposed to TAWL. The only exceptions to this pattern were the kinesthetic and psychomotor ratings during damage assessment.

Table 6

Results of Analyses of Variance Testing for the Effect of Simulation Model in Component Workload Ratings

	Workload Component							
	Visual		Kinesthetic		Cognitive		Psychomotor	
Function	F	p	F	p	F	p	F	p
Verify Weapons	16.01	.0008	11.54	.0032	3.54	NS	11.54	.0032
Designate Target	372.24	.0001	157.05	.0001	195.73	.0001	213.08	.0001
Track/Identify Target	98.59	.0001	15.39	.0001	65.21	.0001	30.97	.0001
Assess Damage	311.20	.0001	295.55	.0001	311.14	.0001	295.55	.0001

Table 7

Workload Component Means and Standard Deviations (in parentheses) for Four Functions from Micro Saint and TAWL

	Workload Component							
	Visual		Kinesthetic		Cognitive		Psychomotor	
Function	Saint	TAWL	Saint	TAWL	Saint	TAWL	Saint	TAWL
Verify	6.25	3.39	0.68	0.26	2.87	2.05	1.49	0.58
Weapons	(0.47)	(3.33)	(0.43)	(0.45)	(2.48)	(2.94)	(0.94)	(0.99)
Designate	5.14	1.23	5.22	1.37	5.90	1.47	4.73	1.20
Target	(0.47)	(2.25)	(1.63)	(2.74)	(1.16)	(2.79)	(0.98)	(2.28)
Track/Identify	8.79	3.80	1.07	0.50	8.90	4.11	1.15	0.48
Target	(2.02)	(5.37)	(1.35)	(1.73)	(3.34)	(5.93)	(1.35)	(1.48)
Assess	6.18	2.17	0.14	0.79	6.01	2.13	0.30	1.75
Damage	(1.18)	(2.40)	(0.20)	(0.41)	(1.14)	(2.32)	(0.43)	(0.89)

DISCUSSION

The results of this investigation have several important implications for workload estimation and task network modeling. First, the outcomes obtained here have demonstrated that the workload estimation technique is sufficiently reliable over time. Practically speaking, therefore, a user who estimates the workload associated with each task in a network would likely produce those same estimates if queried at a later date. Such results help to counteract the subjectivity inherent in this technique. That is, while the exact magnitude of the workload ratings for a task will depend upon the individual who rates the task, the individual's ratings will tend to be consistent over time. Hence, the individual who constructs a model at Time 1 and a modification of the model at Time 2 (e.g., to represent some proposed enhancement or alteration to the system) will most likely apply the workload component rating scales in the same manner each time.

Second, the comparison of two computer modeling tools has revealed that while the estimates of the operator's OW and PW during each function of the mission provided by Micro Saint and TAWL were either similar or identical, the component ratings themselves were not. Of the seven functions that were analyzed; the visual, kinesthetic, cognitive, and psychomotor workload component ratings provided by Micro Saint and TAWL differed significantly on four of them. In most instances, the estimates from Micro Saint exceeded those from TAWL. It is important to point out that the source of this discrepancy most likely is the differential manner in which the half-second transition time between tasks is handled in Micro Saint versus TAWL. In Micro Saint, the transition time was included as part of the completion time for each task; hence, the workload ratings for the task remained in effect through the transition period. In TAWL, on the other hand, the transition is automatically inserted as a blank half-second interval with no workload. These blank intervals would serve to reduce the average workload within each function for TAWL but would leave the general pattern of workload over time unchanged. Along these lines, it should be noted that the overall patterns yielded by the two modeling tools were indeed remarkably similar, as evidenced in part by visual inspection of the plots of operator workload as a function of time for each component (Figures 1 and 2).

In sum, these outcomes imply that if the user's goal is to predict overall and peak workload, either model will suffice. However, if the goal is to conduct a more fine-grained

analysis of component workload estimates, the results will differ depending upon which model is used. If the user decides that the transition times should be regarded as intervals with no workload, then either modeling tool will be appropriate since TAWL's automatic .5 second interval could also be easily implemented in Micro Saint. On the other hand, if the user decides that the transition times might continue to reflect the residual effects of the workload associated with preceding tasks, then the TAWL modeling tool will no longer be an option.

Finally, a comparison of the overall usability of the two modeling tools has indicated that Micro Saint is much more versatile, flexible, and easily utilized than TAWL. The user will be better able to construct a realistic model with the Micro Saint modeling tool. Thus, taken together, the results of this study suggest that the user should seriously consider using Micro Saint rather than TAWL to accomplish task network modeling.

Future Research

Goals for future research include the development of a more realistic Micro Saint model of the F-15E Scud hunt mission. Output from that model will be compared with the workload ratings provided by human operators participating in laboratory and field simulations of the mission. If the predictive validity is sufficiently high, the computer modeling approach will be used further to study other scenarios.

REFERENCES

Bartko, J. J., & Carpenter, Jr., W. T. (1976). On the methods and theory of reliability. The Journal of Nervous and Mental Disease, 163, 307-317.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1987). A comprehensive task analysis of the UH-60 mission with crew workload estimates and preliminary decision rules for developing a UH-60 workload prediction model (Draft Technical Report No. ASI690-302-87[B], Vol. I, II, III, IV). Fort Rucker, AL: Anacapa Sciences, Inc.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1989). Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model: Volume I: Summary report (Research Product 89-08). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ford, C. M., Manton, J. G., & Hughes, P. K. (1990). A worked example of job simulation using Micro Saint (Aircraft Systems Technical Memorandum 125). Melbourne, Victoria, Australia: Defence Science and Technology Organisation Aeronautical Research Laboratory. (AD-A221 479)

Hamilton, D. B., Bierbaum, C. R., & Fulford, L. A. (1991). Task analysis/workload (TAWL) user's guide: Version 4.0 (Research Product 91-11). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD-A241 861)

Hendy, K. C. (Ed.) (1993). Proceedings of the Workshop on Task Network Simulation for Human-Machine System Design. Washington, D. C.: The Technical Cooperation Program, Subgroup U, Technical Panel 7.

Hendy, K. C. (1994a). Survey of national practices in task network simulation for human-machine systems design. Washington, D. C.: The Technical Cooperation Program, Subgroup U, Technical Panel 7.

Hendy, K. C. (1994b). Implementation of a human information processing model for task network simulation (DCIEM No. 94-40). North York, Ontario, Canada: Defence and Civil Institute of Environmental Medicine.

Iavecchia, H. P., Linton, P. M., Bittner, Jr., A. C., & Byers, J. C. (1989). Operator workload in the UH-60A Black Hawk: Crew results vs. TAWL model prediction. In Proceedings of the Human Factors Society 33rd Annual Meeting (pp. 1481-1485). Santa Monica, CA: Human Factors Society.

Jones, A. P., Johnson, L. A., Butler, M. C., & Main, D. S. (1983). Apples and oranges: An empirical comparison of commonly used indices of interrater agreement. Academy of Management Journal, 26, 507-519.

Kameny, I. (Ed.) (1995). Defense Modeling and Simulation Office Data and Repositories Technology Working Group (DRTWG) Meetings Held February 7-10, 1995 and Additional Task Force and Subgroup Meetings Held Between July 1994 and February 1995. RAND National Defense Research Institute.

Kuperman, G. G., Wilson, D. L., & Davis, I. (1993). High resolution radar demonstration program: Operator performance study (Technical Report No. AL-TR-1993-0032). Wright-Patterson Air Force Base, OH: Armstrong Laboratory.

Kuperman, G. G., Wilson, D. L., & Perez, W. A. (1988). Relocatable target acquisition performance with simulated synthetic aperture radar imagery (Technical Report No. AAMRL-TR-88-025). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Laughery, K. R. (1989). Micro Saint: A tool for modeling human performance in systems. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R. Sutton, & L. van Breda (Eds.), Applications of human performance models to system design (pp. 219-230). New York: Plenum Press.

Laughery, R., Drews, C., Archer, R., & Kramme, K. (1986). A Micro Saint simulation analyzing operator workload in a future attack helicopter. Proceedings of the IEEE 1986 National Aerospace and Electronics Conference NAECON 1986, 3, 896-903.

Lawless, M. T., Laughery, K. R., & Persensky, J. J. (1995). Using Micro Saint to predict performance in a nuclear power plant control room: A test of validity and feasibility (Technical Report No. NUREG/CR-6159). Washington, D.C.: Division of Systems Technology, Office of Nuclear Regulatory Research.

McCracken, J. H., & Aldrich, T. B. (1984). Analyses of selected LHX mission functions: Implications for operator workload and system automation goals (Technical Report No. ASI479-024-84). Fort Rucker, AL: U.S. Army Research Institute for the Behavioral and Social Sciences.

Micro Saint [Computer software]. (1992). Boulder, CO: Micro Analysis & Design Simulation Software, Inc.

SAS (The SAS System for Windows 3.10, Release 6.08) [Computer software]. (1992). Cary, NC: SAS Institute Inc.

Siegal, A. I., & Wolf, J. J. (1969). Man-machine simulation models: Psychosocial and performance interaction. New York: Wiley.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Merrill.

GLOSSARY

A/G	Air-to-Ground
CIWAL	Crew-Aiding and Information Warfare Analysis Laboratory
df	degrees of freedom
DMSO	Defense Modeling and Simulation Office
DoD	Department of Defense
FLIR	Forward Looking Infrared Radar
HC	Hand Controller
ICC	Intraclass Correlation Coefficient
INS	Inertial Navigation System
LHX	Light Helicopter
LOCA	Loss of Cooling Accident
Micro Saint	Systems Analysis of Integrated Networks of Tasks for microcomputers
MPD	Multi-Purpose Display
ms	millisecond
MS_B	Mean Square Between groups
MS_W	Mean Square Within groups
nmi	nautical mile
OW	Overall Workload
PB	Push Button
PW	Peak Workload
SAS	Statistical Analysis System
SGTR	Steam Generator Tube Rupture
TAWL	Task Analysis/Workload
TOSS	TAWL Operator Simulation System
WSO	Weapons System Operator